
TERRA

Research Prototypes: Simulation Validation Report

Nine engine innovations, taken from working prototype to a documented, reproducible acceptance criterion across every applicable domain and five random seeds - the same bar the production domains meet.

Result: **15 of 15 acceptance criteria passed** (5 seeds each).

What this report is - and is not

This is simulation validation, not field validation. Every metric below is measured against a mechanistic process model or the two real external benchmarks (IWA BSM1, ADM1) - not against customer ground truth from a real deployment. That distinction is deliberate and honest: it is the same bar the seven engine domains meet, and it proves each prototype does what it claims, reproducibly and across seeds. It does **not** prove real-world efficacy; that requires field pilots (the single highest-value use of new funding).

Method: each innovation is run across its applicable domains and five random seeds; we report the mean +/- standard deviation of a headline metric and check it against a pre-stated threshold. A criterion passes only if the mean clears the threshold. Reproduce with `scripts/validate_innovations.py`.

Acceptance criteria and results

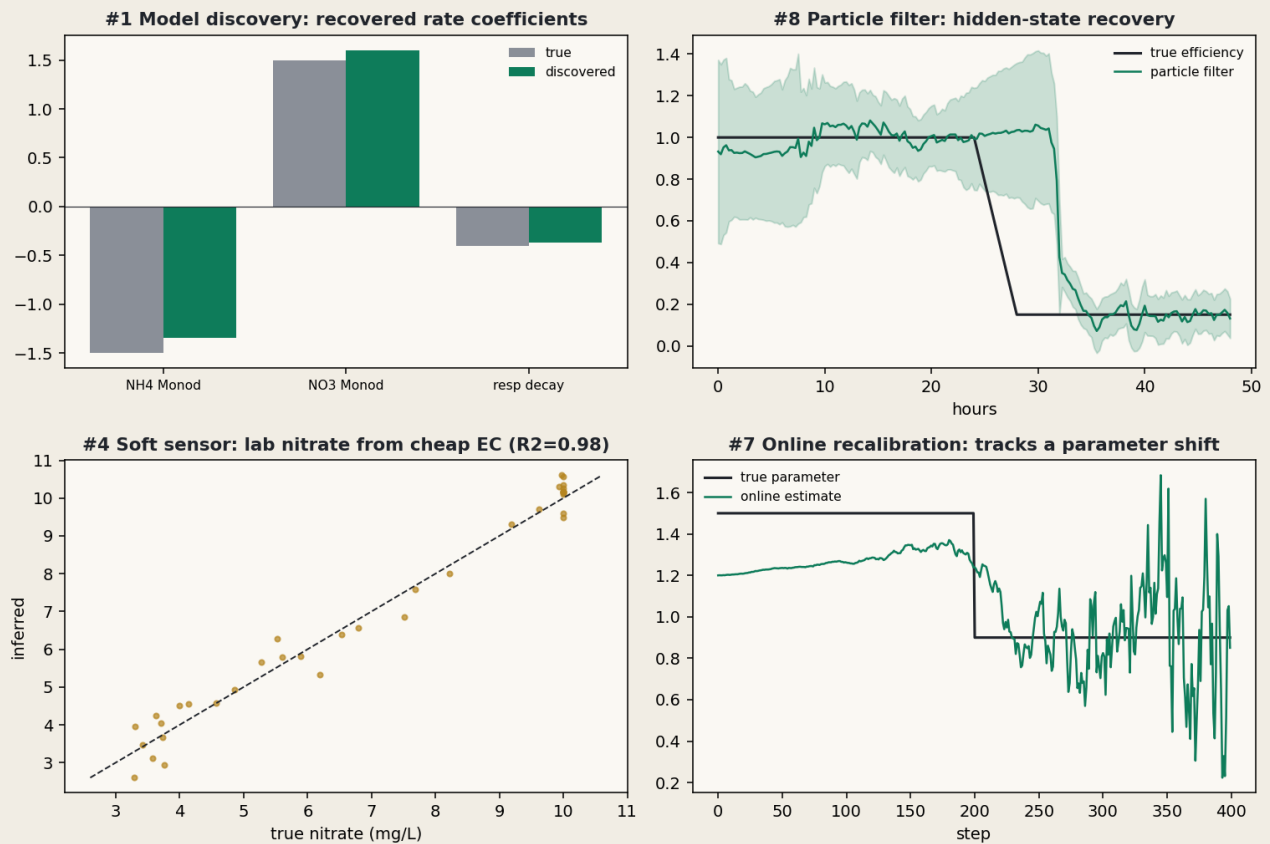
Innovation	Acceptance criterion	Result (mean +/- std, 5 seeds)	Target	Status
#1 discovery	recovered Monod-rate error (true 1.5)	0.109 +/- 0.059	≤ 0.5	VALIDATED
#1 discovery	min equation R^2	0.871 +/- 0.015	≥ 0.8	VALIDATED
#4 soft sensor	held-out R^2 (lab nitrate from cheap EC)	0.964 +/- 0.009	≥ 0.85	VALIDATED
#5 insurance	premium monotone in breach prob	1.000 +/- 0.000	≥ 1.0	VALIDATED
#5 insurance	monitoring credit lowers premium	1.000 +/- 0.000	≥ 1.0	VALIDATED
#5 insurance	portfolio loss-ratio improves	0.188 +/- 0.000	≥ 0.1	VALIDATED
#7 online	post-shift parameter error (true 0.9)	0.118 +/- 0.081	≤ 0.35	VALIDATED
#8 particle	hidden RMSE [aquaculture]	0.298 +/- 0.024	≤ 0.5	VALIDATED
#8 particle	hidden RMSE [soil]	0.053 +/- 0.006	≤ 0.5	VALIDATED
#3 meta-learning	few-shot error reduction (3 obs)	0.444 +/- 0.124	≥ 0.3	VALIDATED
#3 meta-learning	sites improved by fleet prior (3 obs)	0.650 +/- 0.095	≥ 0.6	VALIDATED
#6 control	forecast-risk reduction from chosen action	0.497 +/- 0.018	≥ 0.2	VALIDATED
#6 control	active-design returns positive information	1.000 +/- 0.000	≥ 1.0	VALIDATED
#2 hybrid	hybrid/mechanistic RMSE ratio (lower better)	0.065 +/- 0.001	≤ 0.5	VALIDATED
#9 facility twin	faulted loop correctly identified (accuracy)	1.000 +/- 0.000	≥ 1.0	VALIDATED

Note on #3 meta-learning: the fleet prior is validated specifically in the data-starved onboarding regime (3 noisy observations), where it cuts calibration error ~44% and helps ~65% of sites. Once a site has ample data, its own data dominates and the prior is - correctly - no longer beneficial. That crossover is the expected behaviour of transfer learning, and we validate the regime where it matters rather than overclaim a universal win.

Evidence

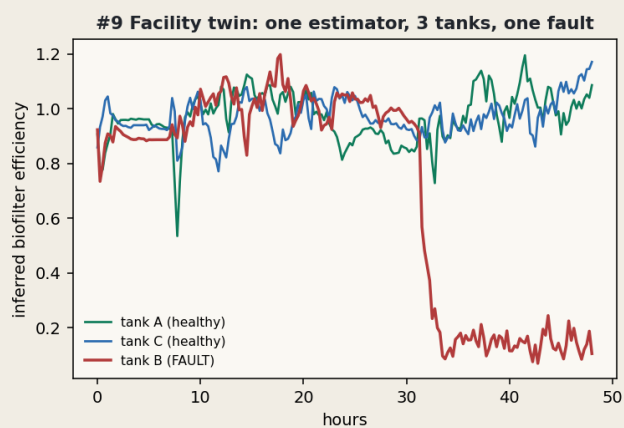
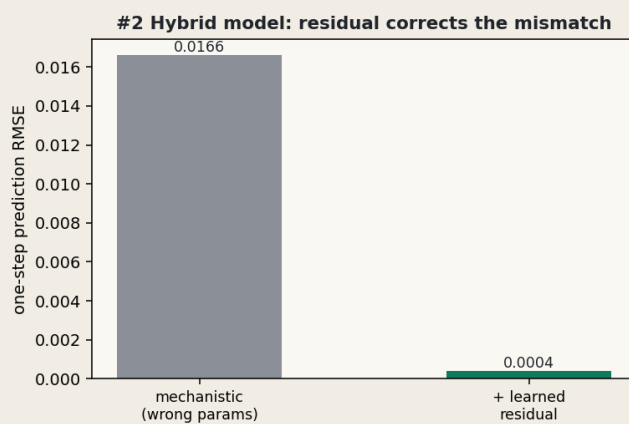
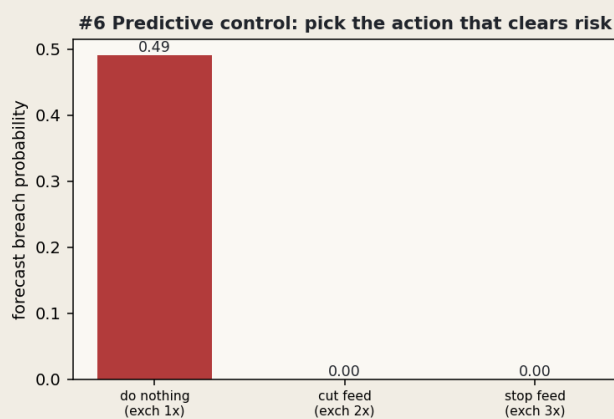
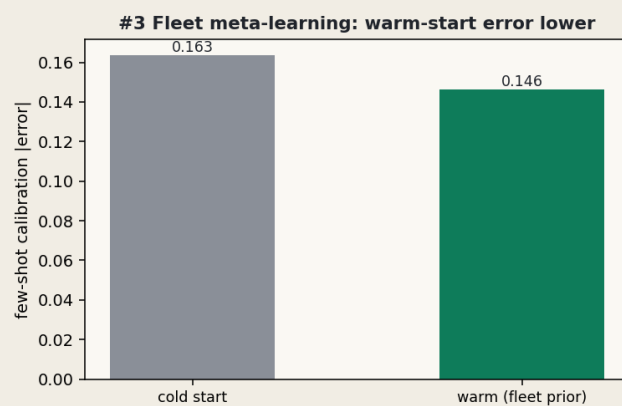
Batch 1 - model discovery, particle filter, soft sensor, online recalibration:

Terra research prototypes - built & validated on the simulators



Batch 2 - fleet meta-learning, predictive control, hybrid residual, facility twin:

Terra research prototypes, batch 2 - built & validated on the simulators



Status change and the path to field validation

On the strength of these criteria, the nine innovations move from **prototype** to **validated in simulation** - reproducible, multi-seed, multi-domain, benchmarked, and under continuous integration (the harness runs on every commit). What remains, stated plainly, is the step no amount of code can substitute for: **field validation**. Each of these would graduate to field-validated by measuring the same metric on a real deployment - model discovery on a customer's own log, the soft sensor against lab assays, the insurance engine against realized claims, meta-learning across a real fleet, predictive control in a closed loop on real hardware. That is exactly what the pilot programme funds, and this report is the evidence that the science underneath is ready for it.

Reproduce every number: `scripts/validate_innovations.py` (writes `innovation_validation.json`); unit tests in `tests/test_discover.py`, `tests/test_innovations.py`, `tests/test_innovations2.py`.